The Collatz Cantor Bijection

Zalán S. Józsa

24/04/2025

Abstract

Starting from the number 1, I build a recursive set—the Collatz TREE—whose branches unfold by inverting the famous Collatz transformation. Within this structure lies an unexpected discovery: a function that encodes any sequence, no matter how long or irregular, into a unique natural number. Each number, in turn, decodes to the same sequence, revealing a perfect symmetry: a bijection between sequences and the natural numbers. This bijection not only confirms the truth of the Collatz Conjecture but shatters long-held assumptions about infinity. Even Cantor's diagonal D—the very cornerstone of uncountability—yields to the method, encoding into a unique number.

Introduction

Georg Cantor, in 1891, formalizes the concept of varying sizes of infinity. His diagonal argument is regarded as the cornerstone of modern set theory, used to argue that $\mathbb{R}$ cannot be put into one-to-one correspondence with $\mathbb{N}$, giving rise to the notion of uncountable sets and the hierarchy of cardinalities.

Some 40 years later Lothar Collatz, in 1937, posed a seemingly simple question: take any positive integer n, and repeatedly apply the transformation $n \mapsto n/2$ if n is even, or $n \mapsto 3n + 1$ if n is odd. His conjecture asserts that this process will eventually reach the number 1, regardless of the starting value. Despite its elementary definition, the conjecture has remained unproven for nearly a century and stood as one of the most well-known open problems in mathematics.

I place Collatz and Cantor in direct opposition—a confrontation between arithmetic and infinity itself. What emerges is decisive: the Collatz Conjecture not only withstands the clash, it utterly dismantles Cantor's hierarchy of the uncountable.

The consequence is unavoidable: accept the Collatz Conjecture—the simplest, most resilient riddle in mathematics—and Cantor's hierarchy collapses; uphold the doctrine of uncountability, and one is forced to deny the conjecture.

The Collatz TREE

I define a recursive set structure, the Collatz TREE, denoted by $\mathcal{T}$, based on powers of two and a conditional branching rule derived from the arithmetic operation $(n-1)/3$.

Let

$$P := \{2^n \mid n \in \mathbb{N}\}$$

be the set of all positive integer powers of 2.

For any $n \in \mathbb{N}$, define the auxiliary set T_n as:

$$T_n := \begin{cases} \varnothing, & \text{if } \frac{n-1}{3} \notin \mathbb{N} \text{ or } \frac{n-1}{3} = 1 \\ \left\{2^k \cdot \left(\frac{n-1}{3}\right) \mid k \in \mathbb{N}\right\}, & \text{if } \frac{n-1}{3} \in \mathbb{N} \end{cases}$$

Define a function $f : \mathbb{N} \to \text{Sets}$ recursively by:

$$f(n) := \begin{cases} n, & \text{if } T_n = \varnothing \\ \{n, \{f(m) \mid m \in T_n\}\}, & \text{if } T_n \neq \varnothing \end{cases}$$

Let

$$\mathcal{T} := \{f(n) \mid n \in P\}$$

That is, $\mathcal{T}$ consists of the recursive expansions generated from each power of 2 in P, beginning with $2^0 = 1$ as the root.

Considerations

Each odd number and its associated sequence of powers of 2 is a branch in $\mathcal{T}$, a more detailed visualization of $\mathcal{T}$ is provided in APPENDIX 1.

The sequence $\{2^k \cdot n\}$ cycles through residues modulo 3 congruent to 1 and 2, and will eventually yield a value m such that $\frac{m-1}{3} \in \mathbb{N}$ and $\frac{m-1}{3} \not\equiv 0 \pmod 3$. This guarantees continued recursive expansion of $\mathcal{T}$.

Assume $\mathcal{T}$ contains a loop. This would require reaching the loop from 1 while being unable to return to 1 (via the standard Collatz transformation). This is a contradiction, therefore, $\mathcal{T}$ cannot contain a loop.

Lemma 1 $\mathcal{T}$ *is loop free.*

To fully make sense of $\mathcal{T}$, an alternative—second visualization is provided in APPENDIX 2.

Collatz Encoder Function

I define a function f that uses a sequence of natural numbers to guide reverse Collatz transformations, starting from 1.

Let $\mathbb{S} := \left(\bigcup_{n\in\mathbb{N}} \mathbb{N}^n\right) \cup \mathbb{N}^{\mathbb{N}}$, the collection of all possible sequences.

Let $s = (a_1, a_2, \dots, a_k) \in \mathbb{S}$ be a sequence.

Let $f : \mathbb{S} \to \mathbb{N}$ be defined by:

$$f(s) := \begin{cases} 0 & \text{if } k = 0 \\ 2^{a_1} & \text{if } k = 1 \\ R_{k-1} & \text{otherwise} \end{cases}$$

I construct the transformation $f(s)$ using the following recursive definitions:

$$R_0 := 1$$

$$R_i := \begin{cases} \text{Process}(a_{i+1}, R_{i-1}) & \text{if } i < k-1 \\ R_{i-1} \cdot 2^{a_{i+1}} & \text{if } i = k-1 \end{cases}$$

$$G_0 := R, \quad G_{i+1} := 2 \cdot G_i$$

$$\mathcal{B}_n(R) := \left\{ \frac{G_i - 1}{3} \;\middle|\; \begin{array}{l} i \geq 1 \\ G_i - 1 \equiv 0 \pmod 3 \\ (i \neq k-2) \Rightarrow \frac{G_i-1}{3} \not\equiv 0 \pmod 3 \\ \frac{G_i-1}{3} \neq 1 \end{array} \right\}$$

$$\text{Process}(n, R) := \text{the } n\text{-th element of } \mathcal{B}_n(R)$$

Explanation

Starting at 1 multiply by 2 repeatedly until reaching a value that is obtainable via $3n+1$ from an odd number not divisible by 3—such that it is the m-th such value where $m =$ the element in s. Apply $(n-1)/3$ to it and move to the next element. Repeat for every element—for the second-to-last element disregard the "not divisible by 3" rule and for the last element simply multiply by 2 repeatedly m-th times where $m =$ the element. This ensures that all possible steps are accounted for. The empty sequence is assigned 0.

Lemma 2 *Every sequence of natural numbers can be encoded using this procedure, and the result is a natural number in* $\mathbb{N}$*.*

Cantor Decoder Function

I define the inverse function f^{-1}, which takes a natural number and constructs a sequence while applying the standard Collatz rules.

Let $f^{-1} : \mathbb{N} \to \mathbb{S}$ be defined by:

$$f^{-1}(n) := \begin{cases} () & \text{if } n = 0 \\ (m) & \text{if } \exists m \in \mathbb{N} \text{ such that } n = 2^m \\ \text{Reverse}(D(n)) & \text{otherwise} \end{cases}$$

Let:

$$n_0 := \min\left\{k \in \mathbb{N} \mid \frac{n}{2^k} \equiv 1 \pmod 2\right\} \quad \text{and let } x_0 := \frac{n}{2^{n_0}}$$

$$d_0 := n_0$$

For each x_i:

$$y_0^{(i)} := 3x_i + 1$$

$$y_{j+1}^{(i)} := \left\{ \frac{y_j^{(i)}}{2} \;\middle|\; y_j^{(i)} \equiv 0 \pmod 2 \text{ and } y_j^{(i)} \neq 1 \right\}$$

I define:

$$d_i := \left| \left\{ j \in \mathbb{N}_{\geq 0} \;\middle|\; \begin{array}{l} y_j^{(i)} \equiv 0 \pmod 2 \\ y_j^{(i)} - 1 \equiv 0 \pmod 3 \\ \left(i = 1 \;\vee\; \frac{y_j^{(i)}-1}{3} \not\equiv 0 \pmod 3 \right) \\ \frac{y_j^{(i)}-1}{3} \neq 1 \end{array} \right\} \right| - 1$$

$$x_{i+1} := \text{the last element in the sequence } \left\{ y_j^{(i)} \right\}$$

Construct:

$$D(n) := (d_0, d_1, \ldots, d_r) \quad \text{such that } x_r = 1$$

$$f^{-1}(n) := \text{Reverse}(D(n))$$

Lemma 3 *Every natural number can be decoded by this procedure, yielding a sequence of natural numbers.*

Encoder-Decoder Inversion

I have shown how to construct a function pair that does the following:

The encoder function f takes a sequence and uses it's elements to apply reverse Collatz transformations to produce an $n \in \mathbb{N}$.

The decoder function f^{-1} performs the reverse: it takes an initial $n \in \mathbb{N}$ and while applying the standard Collatz transformations, produces a sequence.

Together, they satisfy:

$$f^{-1}(f(s)) = s \quad \text{and} \quad f(f^{-1}(n)) = n$$

Lemma 4 *The functions f and f^{-1} are mutual inverses.*

Proof

$\mathcal{T}$ serves both as a visual representation and as a navigational map for the Collatz encoder function. In other words, the Collatz encoder function can be used to navigate $\mathcal{T}$.

As shown in Lemma 1, there can be no collision when navigating via $\mathcal{T}$.

Every sequence can be encoded into a natural number (Lemma 2), and this encoding is via a loop free path (Lemma 1), it follows that each sequence encodes into a unique $n \in \mathbb{N}$.

Since the encoder can encode every sequence $(a_1, a_2, \ldots, a_k)$ into a unique natural number and the decoder is its inverse (Lemma 4), the decoder must decode the corresponding unique sequence $(a_1, a_2, \ldots, a_k)$ for a given $n \in \mathbb{N}$.

Assume, for the sake of contradiction, that there exists an $n \in \mathbb{N}$ not reachable via $\mathcal{T}$. Any such n can be decoded (Lemma 3) into a sequence $(a_1, a_2, \ldots, a_k)$. Encoding this sequence starting from 1 must reproduce the original n (as established above), which contradicts the assumption that it lies outside $\mathcal{T}$.

Therefore, all $n \in \mathbb{N}$ are reachable via $\mathcal{T}$, and $\mathcal{T} = \mathbb{N}$ which proves the Collatz Conjecture.

This proof can be rewritten as follows:

Theorem 1 *The Collatz Conjecture establishes the bijection between $\mathbb{N}$ and the set of all sequences and this bijection, in turn, confirms the conjecture.*

Conclusion

I have shown that a bijection between sequences to natural numbers is not only possible but unavoidable—and this proves every number is part of the Collatz TREE, thus proving the Collatz conjecture.

Simply put: the conjecture reveals the bijection, and the bijection, in turn, confirms the conjecture.

I invite the reader to not take the proof lightly, but to go over it and build it up step by step.

By constructing $\mathcal{T}$, you start to notice that the same structure can be constructed from every number. At every step in the standard Collatz process you simultaneously construct $\mathcal{T}$ backwards—you notice: every structure is part of another structure.

The best I can describe it is: the forward and backward direction becomes one. Shortly, one can't help but notice that they are thinking both backward and forward at the same time. The individual numbers fade into the background, as the entirety of $\mathbb{N}$ is revealed under this new structure.

$\mathcal{T}$ essentially re-arranges the natural numbers in chunks of infinites. This can be clearly seen in APPENDIX 2, which shows an M.C. Escher-esque visualization of $\mathbb{N}$.

The Collatz conjecture forces you to redraw $\mathbb{N}$ to resemble supposedly uncountable infinities. In fact, it allows you to create your own diagonal and claim that there exists an $n \in \mathbb{N}$ that is not within $\mathbb{N}$ ... which would simply be ridiculous.

Once you've built $\mathcal{T}$, from here the function to encode and decode sequences can be cleverly constructed, and once understood, the solution becomes simple and elegant.

For readers who remain skeptical, in APPENDIX 3, all from within set theory, I show the bijection between a subset of $\mathbb{N}$ and the set of real numbers in the interval $(0, 1)$, supposedly of cardinality $\mathfrak{c} = 2^{\aleph_0}$, supposedly $\aleph_1$ under the continuum hypothesis or $\aleph_2$ under other models.

For readers who still remain skeptical, I refer you to APPENDIX 4.

APPENDIX 1 - Visualizing $\mathcal{T}$

The core of $\mathcal{T}$, the core branch, begins at 1 and includes all powers of 2. I denote this branch as t_1:

$$t_1 = \{1, 2, 4, 8, 16, 32, \dots\}$$

At specific points along this branch, new sub-branches are generated. These occur at values n for which $(n-1)/3 \in \mathbb{N}$. Each such value spawns a new branch t_n, which follows the same structure as t_1, beginning from the result of the $\frac{n-1}{3}$ and proceeding via multiplication by 2.

For example, when $16 \in t_1$, it satisfies the branching condition:

$$\frac{16-1}{3} = 5$$

This generates a new branch:

$$t_5 = \{5, 10, 20, 40, 80, \dots\}$$

Since $40 \in t_5$ also satisfies the condition $(40-1)/3 = 13$, another branch, t_{13} emerges:

$$t_5 = \{5, 10, 20, \{40, t_{13}\}, 80, \dots\}, \quad \text{where } t_{13} = \{13, 26, 52, \dots\}$$

We observe that $\mathcal{T}$ contains three distinct types of sets:

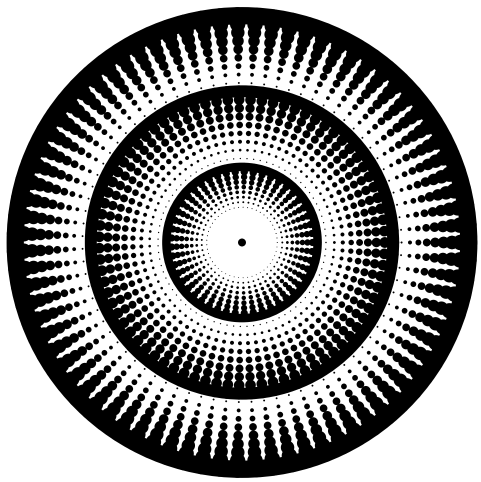
- **Branches with bifurcations** — as defined above, they generate new branches.
$$t_5 = \{5, 10, 20, \{40, t_{13}\}, 80, \dots\}$$
- **Non-bifurcating branches** — numbers divisible by 3 don't satisfy the condition $m = \frac{n-1}{3}$, so they can't generate new branches.
$$t_3 = \{3, 6, 12, 24, 48, \dots\}$$
- **Bifurcations** — sets containing exactly two elements: a number n and a branch t_m, where $m = \frac{n-1}{3}$, I denote it as b_n:
$$b_{40} = \{40, t_{13}\}, \quad b_{52} = \{52, t_{17}\}, \quad b_{10} = \{10, t_3\}$$

$\mathcal{T}$:
{1, 2, 4, 8, b_{16}, 32, ...}
{1, 2, 4, 8, {16, t_5}, 32, ...}
{1, 2, 4, 8, {16, {5, b_{10}, 20, b_{40}, 80, ...}}, 32, ...}
{1, 2, 4, 8, {16, {5, {10, t_{10}}, 20, {40, t_{13}}, 80, ...}}, 32, ...}
{1, 2, 4, 8, {16, {5, {10, {3, 6, ...}}, 20, {40, {13, 26, b_{52}, ...}}, 80, ...}}, 32, ...}

APPENDIX 2 - A second, alternative visualization

Another way to visualize $\mathcal{T}$ is with 1 at the center, and its powers of 2 expanding outward towards infinity—represented by a boundary—the circle.

Similar to M.C. Escher's *Circle Limit* drawings—except after the edge a new set of powers of 2 expand outward towards the next boundary. The first 3 (innermost) rings of the structure, with $n = 1$ at the center:



Inside the inner circle, the center 1; then its powers of 2 reaching towards infinity:

$$\{2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, \dots\}$$

Once over the boundary, the edge itself is the first set of odds—also infinite:

$$\{5, 21, 85, 341, 1365, 5461, 21845, 87381, 349525, 1398101, \dots\}$$

Their own powers of 2 expand into infinity, outwards towards the next boundary:

$$\{10, 20, 40, 42, 80, 84, 160, 168, 170, 320, \dots\}$$

The next set of odds—sitting on the second edge:

$$\{3, 13, 53, 113, 213, 227, 453, 853, 909, 1813, \dots\}$$

Once again, powers of 2 radiate outward, towards the next boundary. This cycle continues indefinitely.

Note

In a standard Collatz path, when applying $3n + 1$, we enter an inner circle and descend a chain of power's of 2 reaching the next boundary—the next odd. Apply $3n + 1$ again to enter the next inner circle and descend a new chain of powers of 2. This descending structure is visible at every step in a Collatz path.

APPENDIX 3 - Bijection from within Set Theory

I define a Pruned $\mathcal{T}$, denoted $\mathcal{T}^*$, with strict trimming constraints: I remove all even and divisible-by-3 values, and truncate each branch to 10 elements.

Let

$$P := \{2^n \mid n \in \mathbb{N}\}$$

be the set of all positive integer powers of 2.

For any $n \in \mathbb{N}$, define the auxiliary set T_n^* as:

$$T_n^* := \begin{cases} \varnothing, & \text{if } \frac{n-1}{3} \notin \mathbb{N} \text{ or } \frac{n-1}{3} = 1 \text{ or } \frac{n-1}{3} \equiv 0 \pmod 3 \\ \left\{2^k \cdot \left(\frac{n-1}{3}\right) \mid k \in \mathbb{N}\right\}, & \text{if } \frac{n-1}{3} \in \mathbb{N} \end{cases}$$

Define a function $f^* : \mathbb{N} \to \text{Sets}$ recursively by:

$$f^*(n) := \begin{cases} \varnothing, & \text{if } T_n^* = \varnothing \text{ or } n \equiv 0 \pmod 2 \\ \{i-1,\, n,\, \{f^*(m) \mid m \in T_n^*\}\}, & \text{if } T_n^* \neq \varnothing \text{ and } n \in \{n_i\}_{i=1}^{10} \end{cases}$$

Let

$$\mathcal{T}^* := \{f^*(n) \mid n \in P\}$$

Let

$$\mathbb{N}^* := \{n \in \mathbb{N} \mid f^*(n) \neq \varnothing\}$$

I define bifurcations as $b_n := \{i, n, t_n\}$, where $i \in \{0, \ldots, 9\}$ is its index in the previous branch, n the actual value, and t_n the next branch. Visually:

$$\{\{0,\, 5,\, \{\{0,\, 13,\, t_{13}\},\, b_{53},\, \ldots\}\},\, \{1,\, 21,\, t_{21}\},\, \{2,\, 85,\, t_{85}\},\, b_{341},\, b_{1365},\, \ldots\}$$

$\mathcal{T}^*$ is a well-defined, countable, recursively constructed set of cardinality $\aleph_0$.

Consider the set of real numbers in the interval $(0, 1)$ supposedly of cardinality $2^{\aleph_0}$, supposedly uncountable.

For a number in the interval $(0, 1)$, each digit determines which indexed bifurcation to take at each step. Thus, a Collatz encoding function will map each such number via a unique path in $\mathcal{T}^*$ to a unique $n \in \mathbb{N}^*$.

Each value of the form $(a_1, a_2, \ldots, a_k)$, has a unique $n \in \mathbb{N}^*$ sitting on an k-th branch in $\mathcal{T}^*$.

Cantor's diagonal D by construction, differs from every number at at least one digit, and this is exactly why it has a unique bifurcation path in $\mathcal{T}^*$ not shared by any other number.

It follows that D has a uniquely corresponding $n \in \mathbb{N}^*$, sitting on an k-th branch in $\mathcal{T}^*$.

APPENDIX 4 - Cantor's Error

The goal of Cantor's argument is to show:

There does not exist a surjective function $f : \mathbb{N} \to (0, 1)$.

The argument proceeds by contradiction:

1. Assume $f : \mathbb{N} \to (0, 1)$ is a surjection.
2. Construct a number $D \in (0, 1)$ such that D differs from $f(n)$ at the n-th digit.
3. Conclude $D \notin \mathrm{Im}(f)$, violating the surjectivity of f.
4. Therefore, such a surjection cannot exist.

However,

under the assumption that f is surjective, the definition of f requires:

$$\exists m \in \mathbb{N} \text{ such that } f(m) = D\,.$$

Cantor's argument claims:

$$\forall n \in \mathbb{N},\ f(n) \neq D\,,$$

which directly contradicts the surjectivity assumption.

Cantor's diagonal method constructs new real numbers, but it does not demonstrate the non-existence of a function f. If f is assumed to be a surjection, then D must already be in its image by definition.

The contradiction Cantor derives does not arise from a genuine inconsistency, but from ignoring what the assumption logically requires—essentially he disregards the assumption he started with.

The interpretation of the diagonal argument as a disproof of countability is logically unsound under strict formal reasoning.

Summary

I have shown that it is logically evident that Cantor's argument is erroneous.

By constructing the Collatz Encoder f—which Cantor himself assumes in his argument—I demonstrate that it does, in fact, encode D into a unique natural number, making the error in his reasoning even more evident.

This shows that the diagonal is not unrepresentable in $\mathbb{N}$ and therefore cannot escape enumeration, and as such the diagonal method is not a demonstration of uncountability but a trivial construction.